Upatras Institutional Repository

- A means to communicate and disseminate institution’s research and educational outcome

- University of Patras O.P. “Education” project
  - Departmental Actions
  - Central Support Actions
  - Repository: “4th Action for Centralized Support of the Educational Process”
DSpace Solution

- Open source
- Clear metadata scheme support (DC)
- Enhanced search capability
- Interoperability: XML and OAI
- Extensible
- “Preservation-ready”
- Unicode
The need for multilingualism

- Contractual need for bilingualism (Greek & English)
  - Interface (now in DSpace 1.3 alpha)
  - Search & Browse
  - Metadata
  - Item Viewing
  - Dynamic switch between languages

- Why not multilingualism?
**I18Ning DSpace Interface**

- **General Approach**
  - Java I18N branch
    - DSpace Java/JSP application model
  - JSTL `fmt`
    - Seamless integration with JSPs
    - Supports 2 or $n$ languages indifferently

- **1st level: Separate text from presentation**
  - Voluminous!

- **2nd level: Separate text from business logic**
  - Hard! (to discover and implement)
Separating text from presentation

1. Substitute every **HTML** word and phrase in JSPs with `<fmt:message key="..."/>` tags

2. Gather all text in a Resource Bundle text file (Messages_en.properties)
   - Key-value pairs

3. Translate the Bundle to any language!
   - May need to pass through `native2ascii` tool first
Example (excerpt from *home.jsp*)

Before:

```html
<table class="miscTable" width="95%" align="center">
  <tr>
    <td class="oddRowEvenCol">
      <h3>Search</h3>
      <p>Enter some text in the box below to search DSpace.</p>
      <p><input type=text name=query size=20>&nbsp;<input type=submit name=submit value="Go"></p>
  </td>
</table>
```

After:

```html
<table class="miscTable" width="95%" align="center">
  <tr>
    <td class="oddRowEvenCol">
      <h3><fmt:message key="home.search1"/></h3>
      <p><fmt:message key="home.search2"/></p>
      <p><input type=text name=query size=20>&nbsp;<input type=submit name=submit value="<fmt:message key="home.search.button"/>"></p>
  </td>
</table>
```
Need to identify text hardcoded in jsp variables, servlets and classes, e.g:

- Location Bar
  - *administer, my dspace…*

- Browse pages
  - the header title changes based on browsing scope

- Input and submit button values written in servlets
  - Select E-Person, ItemMap

- Month names
  - Greek not yet supported in the default java I18N bundle

- Vocabularies
  - Submit Types list
Separating text from business logic (contd.)

**Approach:**

- **Use of Expression Language (EL)**
  - To set EL string variables based on fmt tags
- **DSpace tags parameters now `<fmt:message.../>` values (previously only strings)**
- **Construct arrays of strings for vocabularies**
  - `ListResourceBundle`
- **Use**
  - `LocaleSupport (javax.servlet.jsp.jstl.fmt)` or
  - `BundleSupport (org.apache.taglibs.standard.tag.common.fmt)`

  to “sense” and retrieve current locale
Setting the Locale

- Override browser’s default by submitting a “locale” parameter
  - At any point – dynamic change
- Causes page reload: Context may be lost!
  - Re-post variables along with locale
- May not always work
  - After deletions / additions (exception)
  - Deactivated under admin, tools and submit paths
<c:if test="${param.locale != null}">
<fmt:setLocale value="${param.locale}" scope="session" />
</c:if>

<fmt:setBundle basename="Messages" scope="session"/>
Text stored in PostgreSQL as Unicode (default)
- Lucene tested to work with Greek
- Text extraction tool also works

Search strings over URL:
- URIEncoding="UTF-8" (Tomcat server.xml)

Sorting
- LC_COLLATE = en_US.UTF-8
- LC_CTYPE = en_US.UTF-8
- Only during initdb!
Multilingual Metadata

Storage Layer
- Ready!
- `item.addDC (element, qualifier, lang, value)`

Interface Layer (Submission process)
- Pull-down lang menu for each input
- Use “add more” button
- Types: submit only type code (e.g. 1, 2…) but store multiple text values in every lang
- Languages: submit and store ISO code
- Review process
Item View

- Depending on selected language (not current interface locale)
  - Main title displayed in any case
  - Other elements displayed based on their lang qualifier
  - Elements without a lang qualifier displayed anyway
  - *Item tag now accepts a lang parameter*
“Multilingual” Items, Communities and Collections

- Multilingual Content approach:
  - Different com-col taxonomies (parallel translations)
  - Store items based on their content language
  - Map items between cols when multilingual
    - Add another file in the bundle…
    - …or language independent (e.g. an image)
  - Content language based Search
    - language.iso field now indexed
“Multilingual” Items, Communities and Collections (contd.)

► Pros
  ▪ No need for multilingual col and com names
    • Would require schema change

► Cons
  ▪ Strenuous maintenance
    • Use of Item map tool (authorization)
    • Maintain consistency between collections
Other pieces

- **News**
  - Messages now reside in resource bundles
  - Can be altered by news-edit tool
    - Monolingual only!

- **License**
  - Duplicate text

- **Mails**
  - Duplicate text
  - Parameterized text deeply hardcoded
    - Not yet resolved!
Current and future progress

- HTML text i18N incorporated in DSpace 1.3 alpha
- Now a i18N wiki spin-off has been initiated
  - [http://wiki.dspace.org/i18nSupport](http://wiki.dspace.org/i18nSupport)
- Parameterized keys (Jozsef Marton)
- Idea: Locale to be implemented as a `org.dspace.core.Context` field
  - Independent and globally accessible
- Upatras Institutional Repository (*demo*)